

Serial No.: 09/745,795

REMARKS

Claims 1-8 and 10-20, as amended, remain herein.

Applicant appreciates the statements in the Office Action that claim 20 would be allowable if rewritten in independent form to include all of the limitations of the independent claim(s) from which it depends.

Claim 1 has been amended to recite character joint probability calculating means for calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in said document database. See specification, page 15, lines 4-9, describing characters C_1 , C_2 ... C_{i-1} appearing immediately next to each other.

Claims 2-4, 6, 10, 11, 16 and 17 have been amended to recite calculating a character joint probability that represents a probability of two neighboring characters appearing immediate next to each other. See specification, page 15, lines 4-9 and page 14, last paragraph.

Serial No.: 09/745,795

Minor additional edits have been made to claims 1-8 and 10-17, and claim 9 has been cancelled without prejudice or disclaimer.

Submitted herewith is "Inventor's Brief Description of Prior Art Systems" (as "Attachment A") for clarifying differences between prior art systems and the presently claimed invention.

1. Claims 1, 10 and 16 were rejected under 35 U.S.C. §102(e) over Abe U.S. Patent 6,173,253.

The presently claimed character string dividing system for segmenting a character string into a plurality of words includes input means for receiving a document; document data storing means serving as a document database for storing a received document; character joint probability calculating means for calculating a character joint probability that represents a probability of two neighboring characters appearing immediate next to each other in said document database; probability table storing means for storing a table of calculated character joint probabilities; character string dividing means for segmenting an

Serial No.: 09/745,795

objective character string into a plurality of words with reference to said table of calculated character joint probabilities; and output means for outputting a division result of said objective character string. This system is nowhere disclosed or suggested in the cited reference.

The present invention realizes accurate word division or segmentation without using "tagged corpora" (i.e., text consisting of divided words) or dictionaries. See again Attachment A, briefly describing prior art word dividing or segmenting systems. Contrary to such prior art systems, the word dividing or segmenting system of the presently claimed invention is robust against changes of words or occurrence of new words occurring over a relatively long time, and is applicable to all kinds of documents.

The Office Action cites Abe '253 as allegedly disclosing character joint probability calculating means ("transition probability") for calculating a joint probability of two neighboring characters appearing in the document database. Actually, Abe '253, column 8, lines 30-31, discloses calculation of the transition probability between words, not characters.

Serial No.: 09/745,795

Accordingly Abe '253 does not disclose calculating a character joint probability that represents a probability of two neighboring "characters" appearing immediately next to each other in the document database, as recited in applicant's claim 1.

The system disclosed in Abe '253 relates to input of characters with a pen or a comparable medium. That system provides an automatic function for interpolating (i.e. adding) appropriate supplementary characters in response to the entry of one or more characters. For example, when a Japanese character appears, the Abe '253 system automatically interpolates, or adds, a Japanese character string following such appearing Japanese character. The system of Abe '253 does not calculate the probabilities required for determining a division point between characters. That is, Abe '253 does not disclose (1) character joint probability calculating means for calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in said document database, (2) probability table storing means for storing a table of calculated character joint probabilities,

Serial No.: 09/745,795

and (3) character string dividing means for segmenting an objective character string into a plurality of words with reference to said table of calculated character joint probabilities, as recited in applicant's claim 1 and corresponding recitation of division pattern producing means in claims 10 and 16.

Furthermore, to search an appropriate word with reference to an abbreviated character or characters, Abe '253 considers the connection between an abbreviated word (i.e. candidate word) and characters before and after the abbreviated word. To estimate the connection between appropriate words, Abe '253 uses a method (and means employing such method) for calculating the connection probability of words. This approach is based on the concept that a sentence is plausible when the connection probabilities between words are high. This is a basic principle of the generally known prior art "word model" system.

On the contrary, the presently claimed invention is based on the concept that setting division points where the connection probabilities of characters are low is preferable in obtaining preferable division result of words. In other words, the present

Serial No.: 09/745,795

invention, unlike Abe '253, does not use the connection probabilities between words.

For the foregoing reasons, Abe '253 fails to disclose all elements of applicant's claimed invention, and therefore is not a proper basis for rejection under §102. And, there is no disclosure or teaching in Abe '253 that would have suggested the desirability of modifying any portions thereof effectively to anticipate or suggest applicant's presently claimed invention. Accordingly, reconsideration and withdrawal of this rejection are respectfully requested.

2. Claims 2, 7 and 8 were rejected under 35 U.S.C. §102(e) over Halstead, Jr. et al. U.S. Patent 5,963,893.

The presently claimed character string dividing method for segmenting a character string into a plurality of words includes (1) statistically calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database, and (2) segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities

Serial No.: 09/745,795

so that each division point of said objective character string is present between two neighboring characters having a smaller character joint probability. This method is nowhere disclosed or suggested in the cited reference.

Halstead '893 discloses a method wherein text is divided into words without using dictionaries. Thus, the system of Halstead '893 belongs to the type of word division or segmentation system described as "1.1.3" in Attachment A.

The Office Action cites Halstead '893 as allegedly disclosing calculating a joint probability of two neighboring characters appearing in a given document database. Actually, Halstead '893, column 5, lines 41-55 (especially line 42) describes calculating the probabilities for the breaking unigram (i.e., one character) and bigrams (i.e., two characters) by processing "tagged corpora" (i.e., text consisting of divided words) that identify breaking unigrams and bigrams. Thus, the probabilities of Halstead '893 represent the probability of a word division point appearing when a certain character string appears. Thus, Halstead '893 requires tagged corpora to obtain

Serial No.: 09/745,795

such probabilities, i.e., such system requires previously constructed word-divided texts.

In contrast, the presently claimed invention does not utilize previously devised aides for calculating probabilities. Instead, the probability used in the presently claimed invention is the probability of a character appearing immediately after a certain character (or character string) and accordingly is different from the probabilities disclosed in Halstead '893. The presently claimed invention character string dividing system is different from Halstead '893 because it does not require the preparatory learning operations for obtaining the probabilities with reference to tagged corpora.

Accordingly, Halstead '893 does not disclose (1) statistically calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database, and (2) segmenting an objective character string into a plurality of words with reference to calculated character joint probabilities so that each division point of said objective character string

Serial No.: 09/745,795

is present between two neighboring characters having a smaller character joint probability, as recited in applicant's claim 2.

For the foregoing reasons, Halstead '893 fails to disclose all elements of applicant's claimed invention, and therefore is not a proper basis for rejection under §102. And, there is no disclosure or teaching in Halstead '893 that would have suggested the desirability of modifying any portions thereof effectively to anticipate or suggest applicant's presently claimed invention. Claims 7 and 8, which depend from claim 2, are allowable for the reasons explained herein for claim 2. Accordingly, reconsideration and withdrawal of this rejection are respectfully requested.

3. Claims 4-6 were rejected under 35 U.S.C. §102(e) over Yamamoto et al U.S. Patent 6,098,035.

The presently claimed character string dividing method for segmenting a character string into a plurality of words includes calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database prepared for

Serial No.: 09/745,795

learning purpose. This method is nowhere disclosed or suggested in the cited reference.

The Office Action cites Yamamoto '035 as allegedly disclosing calculating a joint probability of two neighboring characters appearing in a given document database. However, as apparent from the description in Yamamoto '035, column 10, lines 45-46, Yamamoto '035 uses a table of expanded characters. The expanded characters are formed by adding expansion information including at least word division information. In this respect, the system of Yamamoto '035 is based on the "probability of a word division point appearing when a certain character string appears." In obtaining the table (i.e. the relationship between character strings and division points), it is necessary to perform learning operations (as apparent from Yamamoto '035 column 10, line 63). Regarding the operations required for obtaining an expanded character table, Yamamoto '035 performs the learning operations based on a tagged corpus or based on the output results of existing morphological analysis systems as disclosed in Column 13, lines 53-65.

Serial No.: 09/745,795

On the other hand, the probabilities of the presently claimed invention represent the "probability of a character appearing immediately after a certain character (or character string)." Accordingly, the probabilities of the presently claimed invention simply represent a joint probability between a character (or a character string) and another character. The probabilities of the present invention can be calculated from any kind of text, such as a text consisting of non-divided words. Accordingly, the probabilities of the present invention are different from the probabilities of the Yamamoto. The system of this invention requires no learning operations performed based on previously constructed tagged corpora.

Accordingly, Yamamoto '035 does not disclose a character string dividing method for segmenting a character string into a plurality of words that includes calculating a character joint probability that represents a probability of two neighboring characters appearing immediately next to each other in a given document database prepared for learning purpose, as recited in claims 4-6.

Serial No.: 09/745,795

For the foregoing reasons, Yamamoto '035 fails to disclose all elements of applicant's claimed invention, and therefore is not a proper basis for rejection under §102. And, there is no disclosure or teaching in Yamamoto '035 that would have suggested the desirability of modifying any portions thereof effectively to anticipate or suggest applicant's presently claimed invention. Accordingly, reconsideration and withdrawal of this rejection are respectfully requested.

4. Claims 3, 11-15 and 17 were rejected under 35 U.S.C. §103(a) over Halstead '893 and Abe '253.

As explained herein, Halstead '893 discloses a system based on the probability of a word division point appearing when a certain character string appears, and accordingly, Halstead '893 requires learning operations based on previously constructed word-divided texts (i.e., tagged corpora). In contrast, the probabilities of the presently claimed invention represent the probability of a character appearing immediately after a certain character (or character string). Accordingly the probabilities

Serial No.: 09/745,795

of the presently claimed invention are completely different from the probabilities of Halstead '893.

Abe '253 does not provide the deficiencies of Halstead '893 explained herein. Abe '253 refers merely to "characters" as indices of word dictionaries. Abe '253 does not calculate the joint probabilities of characters. Abe, column 8, lines 30-31, describes calculating the transition probability between "words," not characters. Accordingly, the transition probability of Abe '253 is not the joint probability of two neighboring characters, as recited in applicant's claim 3.

For the foregoing reasons, neither Halstead '893 or Abe '253 contains any teaching, suggestion, reason, motivation or incentive that would have led one of ordinary skill in the art to applicant's claimed invention. Nor is there any disclosure or teaching in either of these references that would have suggested the desirability of combining any portions thereof effectively to anticipate or suggest applicant's presently claimed invention. Claims 11-15 and 17 are allowable for the same reasons explained herein for claim 3. Accordingly,

Serial No.: 09/745,795

reconsideration and withdrawal of this rejection are respectfully requested.

5. Claim 9 was rejected under 35 U.S.C. §103(a) over Halstead '893, claim 18 was rejected under 35 U.S.C. §103(a) over Halstead '893, Abe '253 and Yamamoto '035, and claim 19 was rejected under 35 U.S.C. §103(a) over Halstead '893, Abe '253 and Hon et al. U.S. Patent 5,852,801.

Claim 9 has been cancelled, thereby mooted its rejection; claims 18 and 19 depend on claim 17, and are allowable for the reasons explained herein for claim 17.

Moreover, Abe '253 does not provide the deficiencies of Halstead '893 in connection with claim 2, as explained herein. Additionally, Abe '253 and Yamamoto '035 do not provide the deficiencies of Halstead '893 in connection with claims 17 and 18, as explained herein, because they rely on dictionaries and/or tagged corpora. Hon '801, cited for allegedly teaching constant values given to unknown words and to dictionary words, does not teach or disclose applicant's character joint

Serial No.: 09/745,795

probability, and therefore does not provide the deficiencies of Halstead '893 or Abe '253.

For the foregoing reasons, none of Halstead '893, Abe '253 Yamamoto '035 or Hon '801 contains any teaching, suggestion, reason, motivation or incentive that would have led one of ordinary skill in the art to applicant's claimed invention. Nor is there any disclosure or teaching in any of these references that would have suggested the desirability of combining any portions thereof effectively to anticipate or suggest applicant's presently claimed invention. Accordingly, reconsideration and withdrawal of this rejection are respectfully requested.

All claims 1-8 and 10-20 are now proper in form and patentably distinguished over all grounds of rejection stated in the Office Action. Accordingly, allowance of all claims 1-8 and 10-20 is respectfully requested.

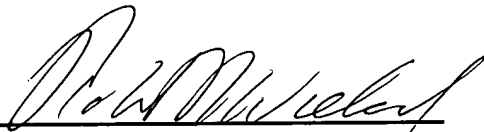
Serial No.: 09/745,795

Should the Examiner deem that any further action by the applicants would be desirable to place this application in even better condition for issue, the Examiner is requested to telephone applicant's undersigned representatives.

Respectfully submitted,

PARKHURST & WENDEL, L.L.P.

October 22, 2004
Date



Roger W. Parkhurst
Registration No. 25,177
Robert N. Wieland
Registration No. 40,225

RWP:RNW/mhs

Attachment: Attachment A "Inventor's Brief
Description of Prior Art Systems"

Attorney Docket No.: PADE:056

PARKHURST & WENDEL, L.L.P.
1421 Prince Street, Suite 210
Alexandria, Virginia 22314-2805
Telephone: (703) 739-0220



ATTACHMENT A

INVENTOR'S BRIEF DESCRIPTION OF PRIOR ART SYSTEMS
SERIAL NO.: 09/745,795 - YASUKI IIZUKA

1. Conventional Prior Art Systems and Their Problems

1.1. Word Dividing or Segmenting Systems are roughly classified into the following four types:

1.1.1 A word dividing or segmenting system relying on only dictionaries

This is a general word dividing or segmenting system, requiring creation of dictionaries. The costs for such dictionaries and their maintenance are very high. The disadvantage of relying on dictionaries resides in the difficulty in identifying a correct path among a plurality of candidate paths.

For example, given a character string of ABCDE, a dictionary may store AB, ABC, CDE, and DE as recognized words. In this case, the following division or segmentation will result from analysis on the dictionary:

AB·CDE

ABC·DE

However, according to this method, there is no reliable way to decide or identify a correct division point between two candidates AB·CDE and ABC·DE. Applicant's specification, page 35, line 30 to page 36, line 7, with reference to Fig. 17A, describes a similar example.

1.1.2. A word dividing or segmenting system relying on both dictionaries and tagged corpora

To solve the problem of the above-described system of 1.1.1, other prior art systems use, as knowledge, the likelihood of connecting a word to another word. To this end, such systems memorize numerous probabilities each representing the likelihood of connecting a word and another word.

According to such system, selection of a correct selection between two candidates AB·CDE and ABC·DE is dependent on memorized probabilities. Namely, a connection probability between AB and CDE is compared with a connection probability between ABC and DE. Then, the candidate having a higher connection probability is selected as a correct choice.

In general, the connection probability between a word and another word is referred to as a "language model" or a "word model." This is also referred to as an "N-gram model" in the case of obtaining the connection probability of a total of N words.

In the case of using the word model, it is necessary to prepare a "tagged corpus" to learn the connection probability between a word and another word. The "tagged corpus" is a bulky text database of numerous words having been divided or segmented beforehand. In this respect, it is mandatory that such system requires the preparation of a "tagged corpus". Details of "tagged corpus" will be explained later.

1.1.3. A word dividing or segmenting system relying on only tagged corpora

The above-described word dividing or segmenting systems 1.1.1 and 1.1.2 are basically dependent on usage of dictionaries. However, preparation and maintenance of such dictionaries is time-consuming. To solve this problem, there is another prior art conventional word dividing or segmenting system that does not rely on dictionaries, as described by Yamamoto '035 and Halstead '893. According to such systems, a text is divided or segmented into words with reference to "tagged corpora (i.e. bulky text database of numerous words having been divided or segmented beforehand)" to check the appearance probabilities of a division point and a character string. In this respect, this system necessarily requires the preparation of "tagged

corpora (i.e. bulky text database of numerous words having been divided or segmented beforehand)."

1.1.4. A word dividing or segmenting system not relying on both dictionaries and tagged corpora

This kind of word dividing or segmenting system is the presently claimed invention.

The following table summarizes the above-described four systems.

		Is preparation of dictionaries necessary?	
		YES	NO
Is preparation of tagged corpora necessary?	YES	System of 1.1.2	System of 1.1.3
	NO	System of 1.1.1	System of 1.1.4

1.2. Problem of Using Dictionaries

In respective languages, regardless of the differences in Japanese, Chinese, and English, on a continuing basis there are new words and changing ways of using certain words. For example, even in a short period of one year, a great number of new words will appear and the way of using those words will change. Furthermore, when the same word appears in different types of documents, such as in medical documents, economic documents and patent documents, the way of using such word may be differentiated according to the type of document. Accordingly, such systems always require substantial periodic maintenance of such dictionaries, and inevitably, a new word will be encountered before it has been added to the dictionary, which degrades the performance of the system.

Inventor's Brief Description of Prior Art Systems
Serial No.: 09/745,795 - Yasuki IIZUK

In this respect, the presently claimed invention can realize accurate word division or segmentation without relying on dictionaries. Accordingly there is no substantial expense relating to dictionary maintenance, and all categories of documents can be processed.

1.3. Problem of Using Tagged Corpora

The above-described word dividing or segmenting systems 1.1.2 and 1.1.3 basically rely on tagged corpora. Accuracy in learning the probabilities is dependent on the volume or scale of prepared tagged corpora, because the probability is a statistical value. Thus, a tagged corpus can be inaccurate and a system utilizing learning based on such inaccurate tagged corpora is not accurate in word division or segmentation. In this respect, a great amount of tagged corpora will be required so that each tagged corpus will be accurate.